View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Universiti Teknologi Malaysia Institutional Repository

ABSTRACT:

Data normalization plays a crucial role in metabolomics to take into account the inevitable variation in sample concentration and the efficiency of sample preparation procedure. The conventional methods such as constant sum normalization (CSN) and probabilistic quotient normalization (PQN) are widely used, but both methods have their own shortcomings. In the current study, a new data normalization method called group aggregating normalization (GAN) is proposed, by which the samples were normalized so that they aggregate close to their group centers in a principal component analysis (PCA) subspace. This is in contrast with CSN and PQN which rely on a constant reference for all samples. The evaluation of GAN method using both simulated and experimental metabolomic data demonstrated that GAN produces more robust model in the subsequent multivariate data analysis, more superior than both CSN and PQN methods. The current study also demonstrated that some of the differential metabolites identified using the CSN or PQN method could be false positives due to improper data normalization.